2015 International Siberian Conference on Control and Communications (SIBCON)

The numerical modeling of a posteriori algorithms for the processing of a random wave impulse sequences

Gyulnara Voskoboinikova
Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia,
Novosibirsk State Technical University, Novosibirsk,
E-mail: gulya@opg.sscc.ru

Marat Khairetdinov
Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia,
Novosibirsk State Technical University, Novosibirsk,
E-mail: marat@opg.sscc.ru

Abstract – A new approach to solving the problem of signal processing is proposed and investigated. It is based on the detection and separation of waveforms of a quasiperiodic random impulse sequences. The solution is obtained by a unifying process of discrete optimization. The efficiency of this approach is illustrated by some numerical experiments.

Keywords – a posteriori algorithms, quasiperiodic random impulse sequences, waveforms, detection, signal/noise ratio, separation, mean-square error, discrete optimization, fractal representation, numerical experiments.

I. INTRODUCTION

One of the main requirements to statement of the problem in the radar-location, hydrolocation, seismolocation, is connected with increasing the accuracy of a sources location. In many problems their coordinates, speeds and types contain in arrival times of waves and in initial wave forms accordingly. On the basis of these measurements at a final stage the location problem as inverse problem is solved. Therefore, it is important to increase the accuracy of estimating the wave parameters in noise. In this paper, a new approach of problem solution is proposed. In comparison to the known methods of statistical data processing, it provides increased accuracy of the measurement of wave arrival times and simultaneous selection of their forms. This approach is based on a posteriori computational algorithms of discrete optimization. The results of numerical experiments on estimating the accuracy and noise immunity of the algorithms are presented.

II. PROBLEM STATEMENT

The problem of estimating unknown parameters of an source is reduced to solving the nonlinear system of equations

$$\hat{\eta} = \eta(\gamma, \theta) + \varepsilon \qquad (1)$$

where $\hat{\eta} = (\hat{n}_1,...,\hat{n}_N)^T$ is the vector of measured travel times of waves, $\eta(\gamma,\theta) = (\hat{n}_1,...,\hat{n}_N)^T$ is the N-dimensional vector of calculated travel times (theoretical hodograph) or the regression function, $\varepsilon = (\varepsilon_1,...,\varepsilon_N)^T$ is the residual vector, $\vec{\theta} = (x, y, z, v, t)^T$ is the m-dimensional vector of estimated parameters, $\gamma = (\gamma_1,...,\gamma_N)$ is the matrix of the sensors coordinates, and N is the sensors number. The space coordinates of the source x, y, z are the parameters to be estimated, v is the velocity characteristic of the medium, and t is the time in the source. The parameters are estimated using information about the distribution of the errors $\varepsilon_i = \hat{\eta}_i(x_i,\theta) - \eta_i(x_i,\theta)$. It is in the limiting case transform to a normal distribution with zero mean and given variances: $E\varepsilon_i = 0$, $E\varepsilon_i\varepsilon_j = \sigma_i^2\delta_{ij}$, $\sigma_i = \sigma(x_i)$, where δ_{ij} is the Kronecker symbol, i=1,..., N.

The solution to equation (1) is reduced to solving the inverse problem. In this case, the accuracy of the solution is in estimating errors of the time vector $\hat{\eta}$ characterized by the variance σ_η^2, the errors $\varepsilon = (\varepsilon_1,...,\varepsilon_N)^T$, and choosing sensor arrangement geometry. The major stages of solving the problem are as follows: 1) detection of waves on the background of external noise, measurement of their travel times, and recovery of their wave forms; 2) solution of the inverse problem of calculating the parameters of source from measurement data. The accuracy of definition of moment of the arrivals of waves can be approximately described standard deviation of an error

$$\sigma^2 \approx \frac{\tau}{2\Delta f \cdot (A/\sigma_n)^2}, \qquad (2)$$

where τ, A are duration and amplitude of wave, Δf is band width, σ_n is standard deviation value of noise. At present, there exist sequential algorithms, allowing "on-line" determination of wave arrival times with minimization errors (2). The sequential approach is oriented to obtaining the fastest, but, in the general case, not optimal solution to the problem. It forms a basis for a family of algorithms to detect the times of changes in the properties of signals [2, 3]. In recent years, wavelet filtration algorithms have been widely used to solve problems of the detection [4, 5, 6].

In addition to the sequential algorithms, there are also a posteriori (off-line) algorithms. In contrast to the former, the latter are oriented to obtaining an optimal solution (solution over all accumulated data). In other words, this approach is potentially more accurate than the successive one. However, its algorithmic implementation involves solving discrete optimization problems using cumbersome calculations. Therefore, most existing off-line technologies for solving such problems have several stages (subproblems): for instance, noise filtering with subsequent solution of the problems of detection, estimation, or decision-making. A key shortcoming of step-by-step data processing is as follows: even in the case of optimal solution of subproblems at each stage the resulting solution may not coincide with the optimal one, because, in the general case, the solution found on the basis of conditional extremums must not coincide with the optimal one. In the present paper, another approach, which has been insufficiently studied as applied to geophysical monitoring, is investigated. Within the framework of this approach, a solution to the problem is found in a unified process of discrete optimization without dividing the problem into stages. The following two types of detection are possible: direct estimation of wave arrival times or simultaneous obtaining of arrival time estimates and wave pulse shapes. In this class of algorithms, we will consider those designed for the processing of sequences that change their properties *quasi-periodically* [7,8]. This means that the time interval between two sequential pulses is bounded from above and below by given constants.

III. A POSTERIORI ALGORITHMS FOR DETERMINING THE PARAMETERS OF WAVE FORMS IN NOISE

In this section, the a posteriori algorithms for solving the problems of detection and separation of waveforms presented by a quasi-periodic sequence and distorted by Gaussian noise are justified. We consider two variants of waveforms in a quasi-periodic sequence, both identical and different. To solve the problem, the following model of data for analysis is proposed. Let the vector components $X=(x_0,...,x_{N-1})\in\Re^N$ form the sequence

$$x_n=\sum_{m=1}^{M}u_{n-n_m}(m),\ n=0,...,N-1,\ \text{где}\ (n_1,...,n_N)\in\Omega_M \quad (3)$$

$$\Omega=\bigcup_{M_{\min}}^{M_{\max}}\Omega_M,$$

$$\Omega_M=\{(n_1,...,n_N)\ |0\le n_1\le T_{\max}-q;\ N-T_{\max}\le n_N\le N-q;\ q\le T_{\min}\le n_m-n_{m-1}\le T_{\max},\ m=2,...,M\ \} \quad (4)$$

Assume that $u_i(m)=0$ if $j\ne 0,...,q-1$, at each m=1,...,M, and $M_{\min}$ and $M_{\max}$ are found from the solution to the systems of inequalities in the definition (4), in which q, $T_{\min}$, and $T_{\max}$ are natural numbers. Also assume that $U_m=(u_0(m),...,u_{q-1}(m))$, m=1,...,M, $w=(U_1,....,U_M)$ and $\eta=(n_1,...,n_N)$. Let $0<\|U_m\|^2<\infty$, m=1,...,M. Then, according to the introduced notation, the vector X depends on the pair of sets η and w having the same number of M elements, that is, $X=X(\eta,w)$. Let the random vector $Y=(y_0,...,y_{N-1})$ be the sum of two independent vectors, $Y=Y(\eta,w)+\mathrm{E}$, where $\mathrm{E}=(e_0,...,e_{N-1})\in\Phi_{X,\sigma^2 I}$, $\sigma^2<\infty$. Here $\Phi_{X,\sigma^2 I}$ denotes normal distribution with the parameters $(0,\sigma^2 I)$.

With allowance for the above, the problem of detection of quasi-periodic sequences of waveforms is in finding, with the observed vector Y, the set η according to which the non-observed vector $X(\eta,w)$ was generated. In this model, components of the vectors Y and X correspond to the observed and non-observed signals, and components of the vector E, to noise. The numbers of the vector components are associated with uniform discrete time. Elements of the set $(n_1,...n_M)$ correspond to the arrival times of waveforms, and the q-dimensional set U_m, m=1,...,M, corresponds to a waveform. The values of $T_{\min}$ and $T_{\max}$ are interpreted as the maximum and minimum intervals between two successive forms.

To solve such problems, the principle of maximum likelihood is used. It has been shown by the authors that noise-immune maximum likelihood detection of a given number of unknown wave forms can be simulated by the following discrete extremal problem:

Problem 1. *Given*: a numerical sequence $Y=(y_0,...,y_{N-1})$ natural numbers q, M, $T_{\min}$, and $T_{\max}$. *Required*: a set $\eta=(n_1,...,n_N)\in\Omega_M$ such that

$$F(n_1,...,n_M)=\sum_{m=1}^{M}\sum_{k=0}^{q-1}y^2_{n_m+k}\to\max$$

In the case that all waveforms are identical, that is, $U_m=U=(u_0,....,u_{q-1})$ for every m=1,...,M, and their number M is not known, the problem of detection of these forms induces the following extremal problem:

Problema 2. *Given*: a numerical sequence $Y=(y_0,...,y_{N-1})$ a vector $U=(u_0,....,u_{q-1})$, and natural numbers $T_{\min}$ and $T_{\max}$. *Required*: a set $(n_1,...,n_N)\in\Omega_M$ and its dimension such that

$$S(n_1,...,n_M)=\sum_{m=1}^{M}\sum_{k=0}^{q-1}u_k(u_k-2y_{n_i+k})\to\min\ . \quad (5)$$

The functions F and S are separable. Therefore, problems 1 and 2 are exactly solved by the same method of dynamic programming, but using different recurrence formulas. Problem 1 is solved in time $O(MN^2)$, and problem 2, in time $O(N^2)$. It's induced by the problem of joint detection and estimation of the recurrent form at an unknown number of recurrences is most difficult. The problem is formulated as follows.

Problem 3. *Given*: a numerical sequence $Y=(y_0,...,y_{N-1})$, natural numbers q, $T_{\min}$, and $T_{\max}$. *Required*: a set $(n_1,...,n_N)\in\Omega_M$ and its dimension such that

$$G(n_1,...,n_M)=\sum_{m=1}^{M}\sum_{j=1}^{M}\sum_{k=0}^{q-1}y_{n_m+k}y_{n_j+k}\to\max$$

Optimal values of components of the sought-for set $\hat{U} = (\hat{u}_0,....,\hat{u}_{q-1})$ corresponding to the waveform are found by the formula

$\hat{u}_k = \frac{1}{\hat{M}} \sum_{m=1}^{\hat{M}} y_{\hat{n}_m + k},\ k = 0,...,q-1,$ where $n_m,\ m = 1,...,\hat{M}$, and $\hat{M}$ are elements of the optimal solution to problem 3.

This *NP* problem is difficult. Hence, in the general case its exact solution cannot be found in polynomial time (if $P \neq NP$). Therefore, approximate algorithms are of interest. One of such heuristic algorithms is proposed in the present paper. The idea of the algorithm is as follows: First, find a solution to problem 1 at M=1 for the initial part of the sequence Y containing $T_{\max}$–$q+1$ elements. Then, using the found value of $\hat{n}_1$, find the set $(y_{\hat{n}_1},...,y_{\hat{n}_1+q-1})$. With this set, solve problem 2 setting $U = (y_{\hat{n}_1},...,y_{\hat{n}_1+q-1})$. Finally, using the found set $(\hat{n}_1,...,\hat{n}_{\hat{M}})$, calculate estimates of components of the vector $\hat{U}$. Taking into account the above approach, we propose a two-stage (locally optimal) algorithm for finding an estimate. Namely, at the first stage a rough estimate of the pulse shape is made. At the second stage, this estimate is refined in the process of solving the problem of joint estimation of the pulse shapes and detection of the pulse beginning times. The essence of the first stage is in solving the problem of verifying the hypotheses. The second stage is designed for solving the problem of estimation, which is reduced to minimization of the additive functional.

To verify the performance of the proposed algorithm and study its accuracy, some numerical experiments were made, with simulation of various waveforms and the same forms and duration and complicated by Gaussian noise of different levels. Real waveforms from explosions and vibration sources recorded earlier and various signal/noise ratios were specified. The generated set $(n_1, ..., n_M)$ of random numbers was used to form a sequence of components of the vector X. According to the adopted model, the sequence of components of the vector Y was synthesized as the sum of the vector X and the Gaussian vector E with the distribution parameters $(0, \sigma^2 I)$. As an example, Fig.1 presents, in graphic form, the results of simultaneous detection and separation of waveforms by the algorithm for solving problem 2. This figure shows: a) the generated model noisy sequence and the sequence found by the algorithm for solving problem 2, b) the results of numerical estimation of errors in the separation of identical waveforms in the quasi-periodic sequence on the background of noise for a signal/noise ratio of 1.25. The arrival times for all separated pulses in the both sequences are plotted on the X-axis, at the beginning of each pulse. The series of numerical experiments has shown that the mean absolute error in estimation of the waveform arrival time is 0,047 s. It is by a factor of 3 smaller than for the wavelet filtration algorithm with a threshold detector used to solve the same problem. To verify the quality of the algorithm for waveform estimation, we used a measure of root-mean-square deviation in the form $\delta_U(M) = 1/q \cdot \sum_{k=0}^{q-1} (u_k - \hat{u}_k)^2$, where $u_k, \hat{u}_k$, k=0,...,q-1 are the given and calculated components of the waveform U. The relative root-mean-square error in the waveform estimation for the data in Fig.1 does not exceed 6 %.

IV. Fractals in a Posteriori Algorithms

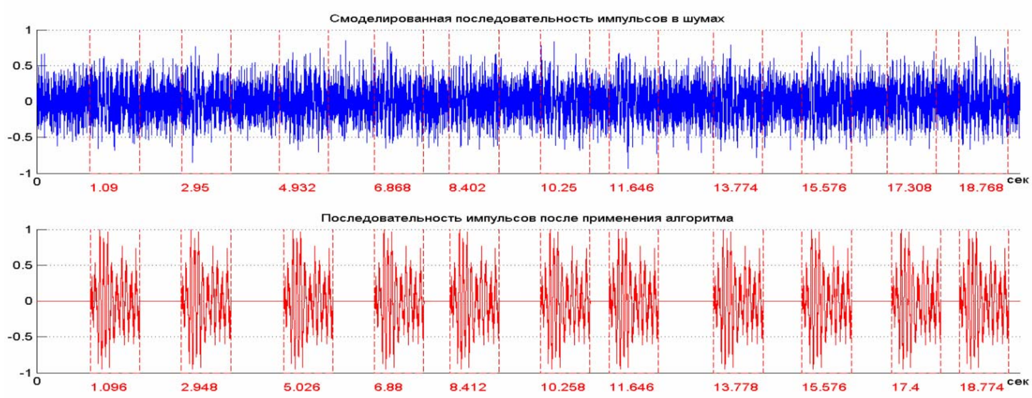
In problems 1-3 and algorithms to solve them, the parameters q, $T_{\min}$, and $T_{\max}$ corresponding to the waveform duration and the upper and lower bounds of the interval between two successive waveforms are input data. However, in practical problems these parameters are often not known in advance.

To remove this a priori indefiniteness, we propose an approach for preliminary estimation of the above parameters based on a fractal representation of waveforms. Waveforms are mapped onto a two-dimensional "frequency-time" plane using a two-dimensional Fourier transform of the form

$$F(k_1, k_2) = \frac{1}{\sqrt{N_1 N_2}} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} F[n_1, n_2] \cdot w_{N_1}^{k_1 n_1} w_{N_2}^{k_2 n_2}, \qquad (6)$$

$$w_N = \exp\left(-j\frac{2\pi}{N}\right)$$

Projection of the function of two variables obtained according to (6) onto the "frequency-time" plane will be a two-dimensional image in which the levels of amplitude values will correspond to brightness levels. The thus obtained waveform images serve for preliminary estimation of the wave pulse boundaries. Subsequent corrected calculation is made using the discrete optimization algorithm by solving problem 3. In what follows, the results of a numerical simulation for the fractal approach to separate the waveform boundaries in noise are presented (see Fig.2).



a)

b)

Fig.1. Signal/noise ratio=1.25, $T_{\min}$=1.3 s, $T_{\max}$=2.2 s, q=1 s; N=20 s, M=11; $\delta_U(M)$=6*10^{-2}.

The simulation was made as follows. Real waveforms taken from experiments were specified. The form corresponding to a specific problem was chosen from the set. Then, a frame was formed with different values of the parameters N, M, $T_{\max}$, $T_{\min}$, and q according to (3), (4). Noise with a Gaussian distribution with the parameters (0, σ) was superimposed on the selected forms. The signal/noise ratio was specified by the level of σ.

Fig. 2 gives a qualitative picture of the above. In Fig.2b (top) one can see noisy waveform sequences to be processed. The record contains 8 waveforms cut out from real seismograms from vibrational sources. Fig.2a presents the result of two-dimensional Fourier transform of the record according to (6). Here the starts and ends of wave pulses, including the beginning of a quasi-periodic pulse sequence, are separated well from noise. This improves the performance of the discrete optimization algorithm. Fig.2b (middle) shows waveform sequences with found arrival times, and in Fig.2c (bottom), the dark histogram shows calculation errors of arrival times without the fractal representation (6), and the light histogram, with the fractal representation. The boundaries in the records outline waveform locations, both initial and calculated ones using the fractal approach with the optimization algorithm. It follows from the error plot that the use of the fractal representation of the pulse sequence allows a considerable increase in the accuracy and reliability of determining arrival times by the discrete optimization algorithm. In some cases, the error decreases by an order of magnitude [11].

V. Solution of the Inverse Problem

The problem of estimating the parameters $\vec{\theta}$ in (1) is a part of regression analysis, and its solution are estimates by the least squares method:

$$\vec{\theta} = \arg\min_{\vec{\theta}\in\Omega} Q(\vec{\theta}),\ Q(\vec{\theta}) = \sum_{i=1}^{N} \sigma_i^{-2}(\hat{n}_i - \eta(\gamma_i,\vec{\theta}))^2 \qquad (7)$$

To find a minimum of the functional $Q(\vec{\theta})$, the Gauss-Newton iterative method or its modifications based on linear approximation of the regression function in the neighborhood of a point $\vec{\theta}^k$ are used:

$$J(\gamma,\vec{\theta}^k)\Delta\vec{\theta}^k + \vec{\eta}(\gamma,\vec{\theta}^k) - \vec{n} + \vec{\varepsilon} = 0 \text{, where} \qquad (8)$$

$$J(\gamma,\vec{\theta}) = \left(\frac{\partial\eta(\gamma_i,\vec{\theta})}{\partial\theta_1}, \frac{\partial\eta(\gamma_i,\vec{\theta})}{\partial\theta_2}, \ldots, \frac{\partial\eta(\gamma_i,\vec{\theta})}{\partial\theta_m}\right), \quad i = 1,2,\ldots,n. \qquad (9)$$

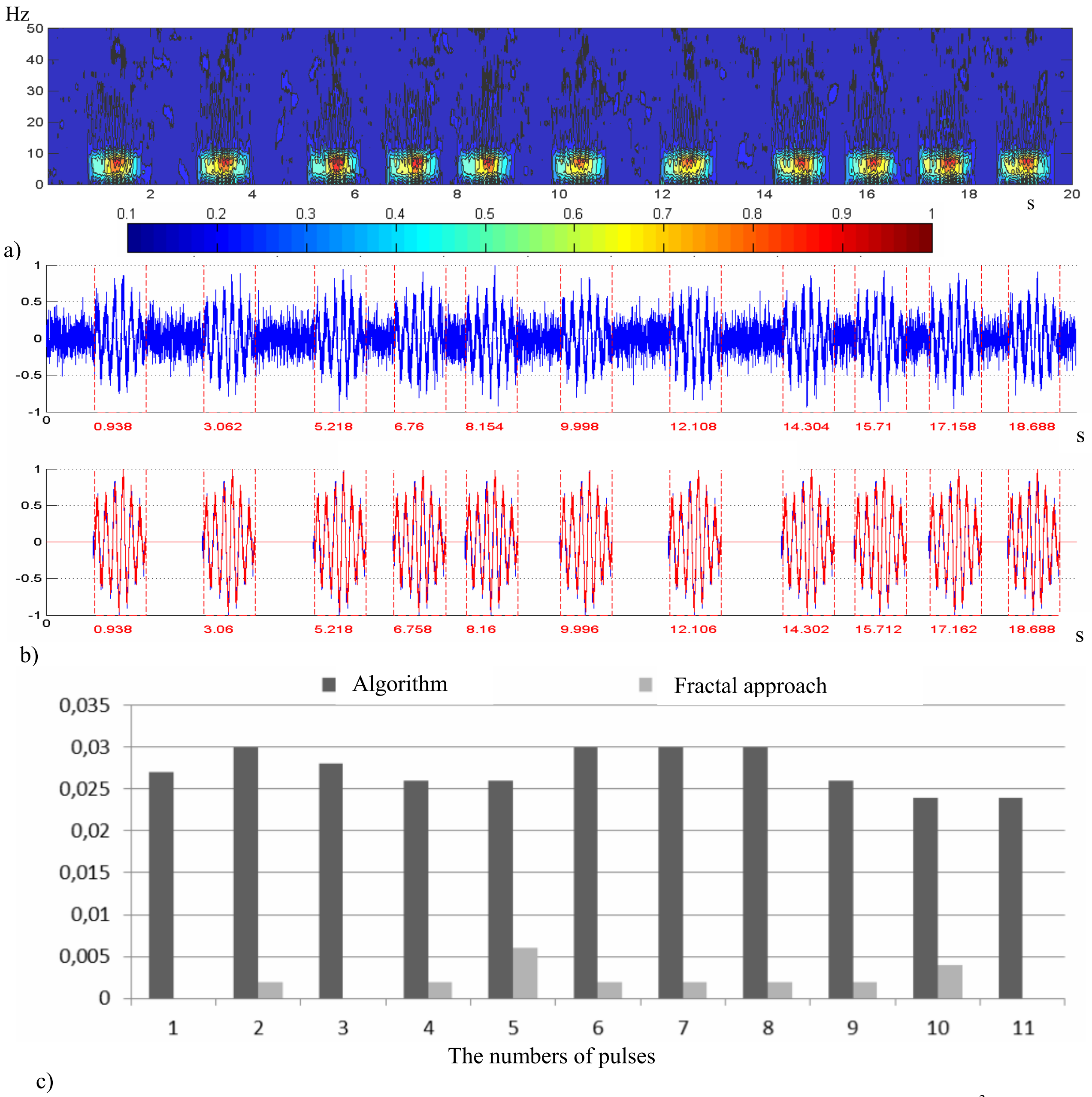


Fig.2. Signal/noise ratio=1.25, $T_{\min}$=1.3 s, $T_{\max}$=2.2 s, q=1 s; N=20 s, M=11; $\delta_M(\sigma)$=2*10^{-3}.

To solve equations (7-9), an approach with direct solution of the system (8) at each step of the iterative process by the pseudo- (or generalized) inversion method is used. It is based on singular value decomposition (*SVD*). It is well-known that the *SSVDC* procedure in the Linpack library is used to calculate *SVD* [10]. The calculation scheme of the SVD procedure is in decomposing the matrix (9) at each step of the iterative process into the product of three matrices

$$J(\gamma,\vec{\theta}^k) = U_k \Sigma_k V_k^T\,, \qquad (10)$$

where U_k is the orthogonal n x n matrix, V_k is the orthogonal m x m matrix, and Σ_k is the diagonal n x m matrix with the structure $\Sigma_k = \begin{pmatrix} S_k \\ 0 \end{pmatrix}$ where $S_k = diag(\rho_1,\ldots,\rho_m)$ is the diagonal matrix of singular numbers arranged in decreasing order $\rho_i \geq \rho_{i+1}$ The method also includes the so-called singular analysis, which is in excluding zero singular numbers and the corresponding columns of the matrices U and V. In this case the iterative process has the following form:

$$\vec{\theta}^{k+1} = \vec{\theta}^k + V_k S_k^{-1} \vec{d}^k,\ k = 0,1,2,\ldots \qquad (11)$$

where $\vec{d}^k$ is a vector consisting of the first m components of the vector $U_k^T \vec{y}(\gamma,\vec{\theta}^k)$, where $\vec{y}(\gamma,\vec{\theta}^k) = (\vec{n} - \eta(\gamma,\vec{\theta}))^T$.

During this process, not only a covariance matrix of the space of parameters, but also a covariance matrix of the space of data, and a resolution matrix $V_kV_k^T$ are obtained. The closeness of the latter matrix to the unit matrix shows the degree of solvability of the problem. The information density matrix is $U_kU_k^T$, whose closeness to the unit matrix shows the relative significance of individual observations [11, 12].

IV. CONCLUSIONS

1. To increase the accuracy of solving the problems of detection and separation of waveforms of a quasiperiodic random impulse sequences a posteriori discrete optimization algorithms have been proposed and analyzed. The solution is obtained by a unifying process of discrete optimization.
2. High accuracy of the proposed algorithms has been proved by numerical experiments. Specifically, it has been shown that the root-mean-square deviation in the estimation of waveforms does not exceed 6 % and relative estimation errors of their arrival times are not worse than 0.1%.
3. The proposed algorithms were used to solve inverse problems of estimation of the parameters of sources, namely, their geographical coordinates and times of occurrence, the velocity characteristic of the medium.

REFERENCES

[1]. *Seismic Tomography*. Edited by G. Nolet, D. Reidel. Publishing Company. Dordrecht, Holland, p.415, 1987.
[2]. *Detection of Abrupt Changes in Signals and Dynamical Systems*. Edited by M. Basseville and A. Benveniste, Springer-Verlag Berlin, Heidelberg, New York, Tokyo, p.278, 1986. .
[3]. Nikiforov I.V. *Successive Detection of changes in the properties of time series*. Moscow: Nauka, 1983.
[4]. Mallat S. A wavelet tour of signal processing. Academic Press, 1999.
[5]. Glinsky B.M., Khairetdinov M.S., Omelchenko O.K., Rodionov Y.I. On a new automized technology for seismic source location. *Bulletin of the Novosibirsk Computing Center, Series: Mathematical Modeling in Geophysics*, iss.8 (2003), p.p. 35-42
[6]. Khairetdinov M.S., Avrorov S.A., Livenets A.A. Computing technology in seismic monitoring networks and systems. *Bulletin of the Novosibirsk Computing Center, Series: Mathematical Modeling in Geophysics*, iss.:13 (2010), p.p.51-69.
[7]. Kel'manov A.V., Jeon B. A posteriori joint detection and discrimination of pulses in a quasiperiodic pulse train // *IEEE Trans. Signal Processing*. 2004. V. 52, N 3. P. 1–12.
[8]. Gruber P. and Todtli J. Estimation of Quasiperiodic Signal parameters by Means of Dynamic Signal Modes // *IEEE Trans. Signal Processing*, vol. 42, No. 3, pp. 552-562, 1994.
[9]. Woskoboynikova G.M. Determination of the arrival times of the seismic by the dynamic programming method // *Proceedings of 9th Korean-Russian International Symposium on "Science & Technology" (KORUS 2005)*. – Novosibirsk: NSTU, 2005. – P. 734-737.
[10]. Kahaner D., MolerC., Nash S. *Numerical Methods and Software*. Prentice Hall, INC. a Pearson Education Company, 1989, p.575.
[11]. Hudson D. *Statistics for Physicists*. Moscow: Mir. 1970. 296 p.
[12]. Forsythe J., Malcolm M., Moler C. *Machine methods of mathematical calculation*. Moscow: Mir. 1980. 280 p.